Bidirectional pre-training method for Neural Network

M. H. Ahmadian
dept. electrical of engineering
University of Ankara
Turkey

V. Sezar
dept. electrical of engineering
University of Ankara
Turkey

A. Mosavian
dept. electrical of engineering
University of Ankara
Turkey

B. Omid
dept. electrical of engineering
University of Texas
USA

Abstract—**In this paper, a two-way pre-training method is presented to combine the learning of deep neural networks with the learning of others. The learning of these networks often does not converge due to facing a high number of local minima. It is now possible to avoid many local minima with the appropriate initial value of network weights. The two-way layer-by-layer pre-training method is a fast and efficient method that adjusts the initial values of its weights in a forward and backward direction using the desired inputs and outputs of the network. For this purpose, a hidden layer based on the weights of the pre-trained layer of the deep network and the auxiliary weights is used to teach the auxiliary networks. Then, the weight values obtained from their training are put under pre-training in the main structure of the network, and for the accurate adjustment of the weights, integral training is done. This method was used for pre-training the weights of three deep neural networks that recognize the person, emotional states and handwritten digits, and it was shown that by using this pre-training method, the speed of learning convergence increases dramatically. Also, the number of recognitions in the face database improves significantly, which indicates the increase in the network's generalization power using this method.**

Keywords- pre-learning, bidirectional, cross-association, deep structure, multi-layer neural networks, convergence of learning

I. Introduction

In the last few years, according to the evidence based on the presence of deep structures in the human brain, researchers have started to learn deep multi-layer neural networks. These structures are capable of learning multiple levels of input. Therefore, it is possible to extract the ranks of the components in them. As the high-level components are formed as a combination of the low-level components in several levels [1]. At present, the usual learning of networks with more than two hidden layers often leads to weak results. So when all the layers are trained through a criterion function such as the degree of similarity to the inputs or the similarity to the desired layers, the results may be worse than shallow models. They depend a lot on the type of functions and initial value of the network parameters [2-4]. In other words, in learning the deep structures of the network, with the increase in the number of layers and neurons, the problem of local minima becomes more noticeable in the middle of the way, it is needed. Therefore, in addition to the direct and usual approach of learning a multilayer neural network that uses the error backpropagation method, there are new approaches that become useful with increasing data complexity and the number of neurons, which are under the heading of advanced methods [6].

Pre-learning methods are used to free the learning process from the topical minims that exist in the middle of the way as a basic obstacle in the learning process. These methods seek to find a suitable starting point for the weights of the network. In addition to facilitating and accelerating the process of learning convergence, pre-learning methods also improve the generalization power of the network. It has been shown in [7-8], that with equal learning error, the network that has an appropriate initial value before learning has a lower test error. In other words, pre-training forces the parameters of the network to be in a region of the parameters space where the desired answer is.

One of the existing methods for pre-training deep networks is the goal-oriented method. This method was used for the first time to learn deep neural networks in [9] to extract the basic non-linear components, and then it was used in [10-11]. In this method, at first, the neural network used is a neural network (BNN), and a hidden layer is assumed, which is added to its depth during training. At each stage of the growth of the number of layers and the depth of the network, the learned values of the weights of the previous shallower network are used as the initial pre-trained values for the deeper network of the next stage [12]. In this way, the initial values of the weights of the deeper neural network are closer to the desired final values than the final values, so they are closer to the final goal. It has been shown that this method can eliminate local minimizations in the path of deep neural networks [13]. In 2006, Hinton presented the method of decomposition into restricted Boltzmann machines for pre-training multilayer neural networks to reduce the nonlinear dimension [14-16]. Other applications have been presented. Considering that the existing pre-training methods have been presented for self-association learning, this method is considered a new solution for the other-association learning of deep neural networks. The two-way layer-by-layer pre-training method, which is a key mode of the layer-by-layer pre-training method [17-18], as a basic pre-training method with the appropriate selection of the initial conditions, helps the error backpropagation algorithm to reach the desired minimum. The presented method for pre-training deep neural networks with the purpose of image recognition has been evaluated.

In continuation, the two-way pre-education method will be introduced. Then, in the third section, the implementations and their results are given. Finally, the summary and conclusion will be presented in the fourth section.

II. Two-way layer-by-layer pre-learning method

Neural networks with deep structures using multiple layers of neurons have the capacity to learn multi-stage and complex nonlinear transformations. But as mentioned, with the increase in the number of neural network layers, it becomes difficult and sometimes impossible to converge their learning through the error backpropagation algorithm due to the existence of local minima. Now, if the number of possible local minimums can be limited in the way of learning the network, the convergence will be improved.

It should be noted that deep neural networks are able to produce outputs with the desired accuracy if they are able to analyze and analyze the input information in different layers without losing information. Therefore, it is necessary to express the different inputs in different layers. If two distinct samples of the input data are reproduced in the network layers with a single and undifferentiated expression from each other, then the nonlinear transformation of the input up to this layer could not work for these samples one by one and discriminatively. Therefore, some of the input information has been lost [2].

In the case of deep neural networks whose neurons use the step function, considering that each of these neurons creates a hypersurface in its input space, they will be able to differentiate the input samples in its two inner layers. Do not separate. In such networks in pattern learning, in order to be able to express all distinct patterns in the output in a different way, it is necessary that all non-matching patterns are expressed with different codes in all layers. For this purpose, it is necessary that in the input space, between the two non-matching samples, at least one hyper-page should pass and the same situation should continue in all the subsequent layers of the network. This means that in all layers, a separate area should be formed by the hypersurfaces of that layer for each layer, so that two separate samples with a single expression are not presented in any layer, and the discriminating information is not reduced. As a result, it means that in the output, the differentiation of the patterns is reduced to a great extent. In the two-way layer-by-layer pre-training method, the above problem can be avoided by placing the neuron surfaces step by step in the appropriate place to preserve the differentiation in the layers. In this method, the desired input and output information is directly used in the initial setting of the weights of the deep network. The initial values of the weights of different layers based on input or output are determined according to which layer is closer. In these stages, due to the fact that the networks of one hidden layer are taught, fast learning and the role of local minimums is less. This pre-training method transfers the initial weights of the deep neural network to a situation that is free from many possible local minima.

In this way, the weights of the neurons of different layers of the network benefit from a relatively efficient pre-training method. It should be kept in mind that this layer-by-layer pre-training method can be used for any convolutional network with any number of hidden layers. It can be shown that the presented method for pre-learning weights is a very efficient method.

In the two-way layer-by-layer pre-training method, if you have a deep network with 2n weight layers according to figure (1), its structure is analyzed in parallel from the beginning and the end, respectively, forward and backward. In this way, in the sequential path, BNNs of a hidden layer is defined whose input weights (Wi) represent the weights of the layer in question from the deep network. Their output weights (Vi) are also auxiliary weights that are used to determine Wis. In this path, the first BNN is trained with the input vector and the subsequent BNNs are trained with the image of the input vector of the previous BNN in their bottleneck layer. In this path, the first BNN is trained with the output vector and the subsequent BNNs are trained with the image of the input vector of the previous BNN in its bottleneck layer. Unlike the BNNs of the forward path, in the backward path, the output weights of the BNNs will be used in the deep network structure.

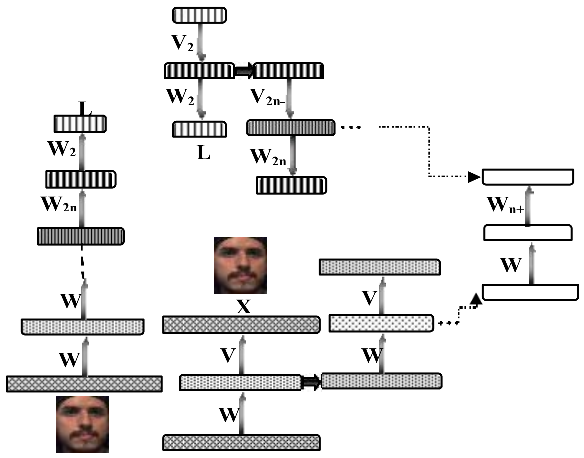

Figure 1. How to analyze a deep neural network in the two-way layer-by-layer pre-training method

For pre-training with the two-way layer-by-layer method to determine the values of weights in the forward path, the local cost functions are defined based on the correct reconstruction of the input of the deep neural network and in the backward path based on the desired output reconstruction [4].

$$E_i^{LF} = \frac{1}{2}\|Y_{i-1}^F - Z_i^F\|^2 \qquad n-1 \geq i > 1 \tag{1}$$

$$E_j^{LB} = \frac{1}{2}\left\|Y_{j-1}^B - Z_j^B\right\|^2 \qquad 2n-1 \geq j \geq n+1 \tag{2}$$

E_i^{LF} and E_j^{BF} in relations (1) and (2), respectively, the self-adherence errors of BNNs i-th and j-th are in the forward and backward paths, which are defined as the input reconstruction error in the output layer. In these relations, Y_{i-1}^F and Y_{j-1}^B the input images of the (i-1)-th and j-th BNNs are in the two paths corresponding to the relations (3) and (4) in their throat layer [6].

$$Y_{i-1}^F = f(Y_{i-1}^F w_{i-1} - b) \qquad n \geq i \geq 2 \qquad (3)$$

$$Y_{j-1}^B = f\left(Y_j^B V_j - b\right) \qquad 2n \geq j \geq n+2 \qquad (4)$$

Also Y_0^F The input vector of the first BNN in the forward path is equal to X, the input vector of the deep neural network, and Y_{2n}^B, the input vector of the first BNN in the backward path, is equal to the desired output vector L of the deep neural network. Z_i^F and Z_j^B also, which are the i-th and j-th BNN outputs in the forward and backward directions, are defined in relations (5) and (6).

$$Z_i^F = f(Y_i^F V_i - b) \qquad n-1 \geq i \geq 1 \qquad (5)$$

$$Z_j^B = f\left(Y_{j-1}^B w_j - b\right) \qquad 2n+1 \geq j \geq n+1 \qquad (6)$$

In the above relationships, b determines the value of the threshold level for the activity function of neurons. This process continues until the n-1st weights from the forward path and the n+2nd weight from the backward path. Then, the images of the inputs in the throat layer of the last two BNNs are used for the calculation of two paths and for the purpose of training a hidden layer consisting of the weights of W_n and W_{n+1}. In this way, the initial values of all the weights of the deep network are obtained, and their learning is based on the error backpropagation algorithm.

After completing the pre-training steps, the obtained weights are considered as the initial weights in the structure of the integrated deep neural network, and with the help of error backpropagation operations, more accurate values of the network's weight matrices are obtained.

III. IMPLEMENTATIONS AND RESULTS

In this section, the two-way layer-by-layer pre-training method is evaluated in order to properly adjust the weights of other deep neural networks. Therefore, its effectiveness in learning deep classification neural networks with the purpose of recognition in several different databases is investigated. Therefore, in the following, three databases will be used.

A. Databases

A.I. Bosphorus database

The Bosphorus database [16] includes 3D and 2D images of the faces of 150 subjects, which were collected at Begaze University, Turkey. This database includes a rich set of facial expressions, head rotation, and different types of obstructions, the collection related to different facial expressions has been used for evaluation in this article. Figure (2) These situations were randomly selected for one of the pictures of 95 people from this group. These images were used in black and white with 256 gray levels. Also, in order to reduce the load of calculations, the clarity of these images was reduced to 114*92.

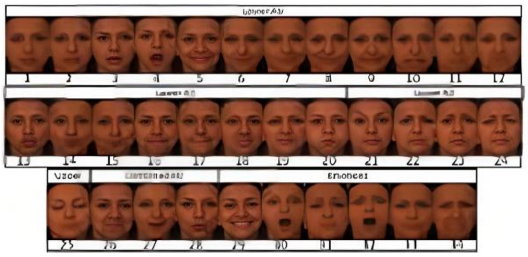

Figure 2. Images related to different states of the face for one of the people from Besphorus database [16]

The pictures of 95 people from this group were randomly selected for examination. These images were used in black and white with 256 gray levels. Also, in order to reduce the load of calculations, the clarity of these images was reduced to 114*92.

In this research, this database has been used for face recognition. Since the number of images related to different states of the face is not the same for all individuals, the images related to 12 common states among all individuals were set aside for training and other images for testing. In the end, the training and test sets included 1126 and 1443 images, respectively. People's label was also defined as 7 binary codes that represent 95 different codes that identify people's identity.

A.II. CK+ database

The CK+ database is an extension of the Cohen-Kund database. [17] In the Cohen-Kund database, there are sequences of images for 6 emotional states of people, which change from the neutral state to the intended state. These emotional states include happy, surprised, angry, scared, hating and upset states. Also, for each face image, there is a vector of signs that can be used to align the faces. In this article, the faces are aligned so that the eyes are in a horizontal direction. Additional information around the images has also been removed in such a way that all the important details are present in the image. Finally, the dimensions of all images were changed to 50x50. In figure (3), some samples of the images of this database are given for the used emotional states.

Figure 3. Samples of Cohen-Kand database images for different emotional states after aligning the faces and removing additional information around the images

Also, the images of the first half of each of these categories were deleted, so that the images that really represent the related emotional state are kept. In this article, this is the base The data has been used to recognize emotional states. Therefore, the images are divided into two groups. The training set is the images related to 96 people (2816 images) that are used for training the deep neural network of the recognizer. The pictures related to the remaining 10 people (351 pictures) are also used for the test. Emotional states are also defined as 4 binary codes that represent the state of people's faces.

A.III. MNIST database

This database includes English handwritten numbers from 0 to 9, the image of each number is stored as a 28x28 pixel gray image. Its training set includes 60,000 images and the test set contains 10,000 images [18]. The labels defined in this article are also 3-way binary codes expressing numbers.

B. Preparation

Each image is considered a two-dimensional array of M×N light values. But since in the input of the network, this array must be presented as one dimension, each image is converted into an MN×1 vector. Also, the pixel values of each face image are normalized to the range of zero to one.

C. Model

The structure of the deep classification neural network for recognition in three databases is given in table (1). This network is a network with three hidden layers, whose input corresponds to the light values of the image pixels and whose output corresponds to the defined labels. This network is trained separately for each data set and its results are evaluated. To show the effectiveness of the two-way layer-by-layer pre-training method, first random initial values are set for the weights, and training is done. In the next step, using the proposed pre-training method, the initial values of the weights are determined and then the network is trained.

TABLE I. PARAMETERS OF DEEP NEURAL NETWORK

The number of hidden layers	3
The number of hidden layer neurons	50-200-400
The function of neurons in the hidden layer	Unipolar nonlinearity
The function of output layer neurons	Unipolar nonlinearity
Learning coefficient	0.001
Torque coefficient	0.7

D. Two-way layer-by-layer pre-training method for deep recognition network pre-training

According to the method presented in section (2), this deep network is broken into two single-layer BNN and another network with one hidden layer. One of the BNNs is trained with a vector of pixels and another with a vector of binary codes defined as corresponding labels. The number of training repetitions in each pre-training stage is assumed to be 500 repetitions for face recognition including Bosphorus and CK+ database images, and 10 repetitions for handwritten recognition due to the large volume of recognition. Because learning more than the size in the pre-learning stages leads to setting more than the limit of the weights on the training judges, as a result of which the generalization power of the deep network will decrease. Then, the images of the inputs are calculated in the hidden layer of each BNN and used to train another network of a hidden layer. In this way, the initial values of all the weights of the deep network are obtained.

E. Results

The graph of Figure 4 shows the changes of learning error for two deep person recognition networks, one with random initial value and the other with initial value by the proposed pre-training method, on Besphorus database. As it can be seen, with the appropriate initial setting of network weights instead of setting random values for them, the speed of learning convergence increases dramatically.

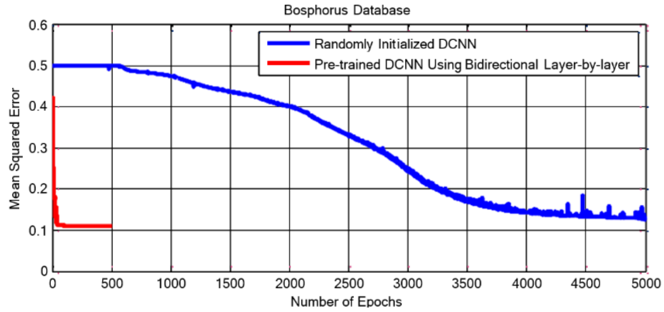

Figure 4. The diagram of the changes of the learning error of the deep network of the person recognizer with two methods of randomly estimating the initial weights and using two-way layer-by-layer pre-training on the Besphorus database

As it reaches its minimum value in the same initial repetitions. For the network with a random value (according to the process of reducing the learning error), the weights of the last iteration of learning and for the pre-trained network, the weights of the 40th iteration (after that, the learning error remains constant). For the recognition of people, the test data was used.

TABLE II. COMPARISON OF THE ACCURACY PERCENTAGE OF PEOPLE RECOGNITION IN BOSPHORUS DATABASE BY THREE HIDDEN LAYERS NEURAL NETWORK AFTER ACCURATE ADJUSTMENT OF WEIGHTS

How to estimate the initial weights of deep neural network	**The number of repetitions**	**Training data**	**Test data**
Randomly	5000	99.72%	86.52%
Two-way layer by layer pre-education	40	100%	97.16%

The results of table (2) show that by using this pre-training method, the results of the person's recognition for the judges of the exam have improved by 11%. It is Especially, in the pre-training stages, one-layer hidden networks are taught, and the duration of their training is much less in each repetition compared to three-layer hidden networks. In the diagram of figure (5) and table (3), the results on the CK+ database with the aim of recognizing the emotional state in face images are given.

As can be seen in the diagram of figure (5), the network error diagram with random initial values does not decrease significantly after the 2500th repetition of learning. Therefore, the weights of this stage are used for the state recognition network. For the pre-trained network, the weights of the 90th repetition of training have been used.

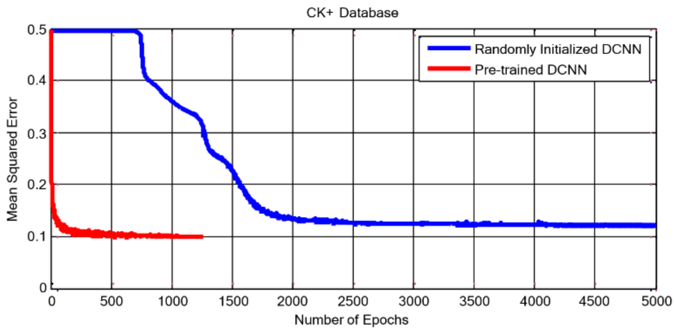

Figure 5. The diagram of the changes of the learning error of the deep network of the person recognizer with two methods of randomly estimating the initial weights and using two-way layer-by-layer pre-training on the CK+ database

The recognition results in table (3) show that the use of the bidirectional layer-by-layer pre-training method for the initial estimation of the weights of the recognition network improves the recognition accuracy percentage by 14%.

TABLE III. COMPARISON OF THE ACCURACY PERCENTAGE OF PEOPLE RECOGNITION IN CK+ DATABASE BY THREE HIDDEN LAYERS NEURAL NETWORK AFTER ACCURATE ADJUSTMENT OF WEIGHTS

How to estimate the initial weights of deep neural network	**The number of repetitions**	**Training data**	**Test data**
Randomly	2500	99.41%	73.70%
Two-way layer-by-layer pre-education	90	99.96%	87.29%

In the case of the MNIST handwritten digit database, due to a large number of samples, the number of repetitions of learning in the pre-learning stages was considered to be 10 repetitions. Experiments also showed that after this number, the error in the pre-training stages for this data does not improve. In the graph of figure (6), the changes of the learning error of the deep network of the digit recognizer in the stage of accurate adjustment of the weights, for two methods of initial estimation of the weights randomly and using the pre-training of the two-dimensional layer are given.

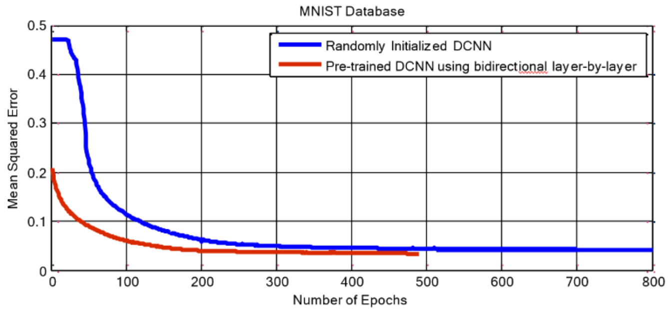

Figure 6. The diagram of the changes of the learning error of the deep network of the person recognizer with two methods of randomly estimating the initial weights and using two-way layer-by-layer pre-training on the MNIST database

Table (4) shows the number recognition error for the two methods of initial estimation of network weights. For handwritten judges, no improvement in digit recognition has been achieved by the pre-trained network. It can be mentioned that the problem of local minima in the learning process of deep networks appears more recently in complex problems such as face and face mode, so the use of appropriate pre-training in their learning process is more effective. But for handwritten numbers, due to the simplicity of the learning task, the estimated network with initial random weights has a favorable result, as the pre-training could not create a significant improvement in the results. Of course, it should be mentioned that the result obtained from the network with pre-training is three times faster. Especially that the stages of its pre-education were short.

TABLE IV. COMPARISON OF THE ACCURACY PERCENTAGE OF PEOPLE RECOGNITION IN MNIST DATABASE BY THREE HIDDEN LAYERS NEURAL NETWORK AFTER ACCURATE ADJUSTMENT OF WEIGHTS

How to estimate the initial weights of deep neural network	**The number of repetitions**	**Training data**	**Test data**
Randomly	600	99.88%	98.24%
Two-way layer-by-layer pre-education	200	99.92%	98.56%

IV. CONCLUSION

In some cases, networks with one or two hidden layers in the input data mapping to the output space do not work successfully. Because for more complex data with high dimension, such as face images, the sudden transformation of the input data features into the output, which can be labeled by the input layers, increases the network error significantly. Therefore, there is a need for deeper structures of neural networks in order to extract data through multiple layers of deeper components, and then with their help, proper mapping is done to the output. This is when the number of layers of neural networks increases, the convergence of their learning faces a problem due to the increase of local minima. In order to solve this problem, some methods for pre-learning the deep structures of neural networks in self-learning applications have been presented so far. However, in this article, the two-way layer-by-layer pre-training method for the appropriate estimation of the initial weights of the multi-layer networks with the other function has been presented, which pre-trains the network weights in parallel in a two-way path forward and backward. This method is presented for other applications of deep neural networks, unlike previous pre-training methods that were defined only for self-associative neural networks. This method was used for pre-training the weights of three deep neural networks for person recognition, classification of emotional states and recognition of handwritten digits, and it was shown that by using this pre-training method, the convergence speed of the learner can be increased. Also, the rate of recognition in the face database improves significantly, which indicates the increase in the generalization power of the network using this pre-training method.

REFERENCES

[1] Ocak, C., "Design, Analysis and Application of a New Three Level Brushless DC Motor for Electric Vehicles", Ph. D. Thesis, Graduate School of Natural and Applied Sciences, Gazi University, Ankara, 2013.

[2] Tousizadeh, M., Che, H.S., Selvaraj, J., Rahim, N.A., Ooi, B. T., "Performance Comparison of Fault-Tolerant Three-Phase Induction Motor Drives Considering Current and Voltage Limits", IEEE Transactions on Industrial Electronics, Vol. 66, No. 4, pp. 2639-2648, April 2019.

[3] Nour, M., Said, S.M., Ali, A., and Farkas, C., "Smart Charging of Electric Vehicles According to Electricity Price", International Conference on Innovative Trends in Computer Engineering, Aswan, Egypt, 2019.

[4] Hori, Y., "Future Vehicle Driven by Electricity and Control-Research on Four-Wheel- Motored "UOT March II", IEEE Transaction on Industrial Electronics, 51(5), pp. 954-962, 2004.

[5] A. Mousaei and M. B. B. Sharifian, "Design and optimization of a linear induction motor with hybrid secondary for textile applications," 2020 28th Iranian Conference on Electrical Engineering (ICEE), 2020, pp. 1-6, doi: 10.1109/ICEE50131.2020.9260773.

[6] Ünlü, N., Karahan, Ş., and Tür, O., "Electric Vehicles", Energy Systems and Environmental Research Institute, 6-22, pp. 42-100, 2003.

[7] Dorrell, D.G., Popescu, M., Evans, L., Staton, D.A., and Knight, A.M., "Comparison of Permanent Magnet Drive Motor with a Cage Induction Motor Design for a Hybrid Electric Vehicle", 2010 International Power Electronics Conference, pp. 1807-1813, 2010.

[8] Grilo, N., Sousa, D.M., and Roque, A., "AC motors for application in a commercial electrical vehicle: Designing aspects", 16th IEEE Mediterranean Electrotechnical Conference, pp.277-280, March 2012.

[9] Ehsani, M., Gao, Y., and Gay, S., "Characterization of electric motor drives for traction applications", The 29th Annual Conference of the Industrial Electronics Society, pp. 891-896, 2003.

[10] Nanda, G., and Kar, N.C., "A survey and comparison of characteristics of motor drives used in electric vehicles", Canadian Conference on Electrical and Computer Engineering, pp. 811-814, 2006.

[11] Tiecheng, W., Ping, Z., Qianfan, Z., and Shukang, C., "Design Characteristic of the Induction Motor Used for Hybrid Electric Vehicles", 12th Symposium on Electromagnetic Launch Technology, pp. 523-527, 2005.

[12] Damiano, A., Gatto, G., Marongiu, I., Porru, M., Serpi, A., "Real-time control strategy of energy storage systems for renewable energy sources exploitation," IEEE Transactions on Sustainable Energy, vol. 5, no. 2, pp. 567-576, 2014.

[13] A. Mousaei, M. B. Bannae Sharifian and N. Rostami, "Direct Thrust Force Control (DTFC) of Optimized Linear Induction Motor with Super Twisting Sliding Mode Controller (STSMC)," 2021 12th Power Electronics, Drive Systems, and Technologies Conference (PEDSTC), 2021, pp. 1-5, doi: 10.1109/PEDSTC52094.2021.9405903.

[14] Mousaei, A. (2023). A Semi-Intelligence Algorithm For Charging Electric Vehicles. *arXiv preprint arXiv:2302.13150*.

[15] Xue, X. D., Cheng, K. W. E. and Cheung, N. C., "Selection of electric motor drives for electric vehicles", Australasian Universities Power Engineering Conference, Hong Kong, pp. 170-175, 2008.

[16] Guzinski, J. and Abu-Rub, H., "Sensorless induction motor drive for electric vehicle application", International Journal of Engineering, Science and Technology, 2 (10), pp. 20-34, 2010.

[17] A. Mousaei, M. B. Bannae Sharifian and N. Rostami, "An Improved Predictive Current Control Strategy of Linear Induction Motor Based on Ultra-Local Model and Extended State Observer," 2022 13th Power Electronics, Drive Systems, and Technologies Conference (PEDSTC), 2022, pp. 12-18, doi: 10.1109/PEDSTC53976.2022.9767535.

[18] Kim, K.T., Song, H.E., and Park, G.S., "A study on the design of induction motor in low speed urban electric vehicle", IEEE Transportation Electrification Conference and Expo, pp. 1-4, Korea, 2016.

[19] Li, K., Cheng, G., Sun, X., Yang, Z., Fan, Y., "Performance optimization design and analysis of bearingless induction motor with different magnetic slot wedges", Results in Physics, 12, pp. 349-356, 2019.

[20] Yahaya, E. A., Omokhafe, T., Agbachi, E. O., James, A. G., "Advantage of Double Cage Rotor over Single Cage Rotor Induction Motor", Innovative Systems Design and Engineering, Vol.6, No.12, pp. 1-4, 2015.

[21] Zhou, G. Y., Shen, J. X., "Current Harmonics in Induction Machine with Closed-Slot Rotor", IEEE Transactions on Industry Applications, Vol. 53, No. 1, pp. 134- 142, 2017.

[22] Gyftakis, K. N., Kappatou, J., "The Impact of the Rotor Slot Number on the Behaviour of the Induction Motor", Advances in Power Electronics, Vol. 2013, pp. 1-9, 2013.

[23] Mousaei, A. (2023). Improving Energy Management of Hybrid Electric Vehicles by Considering Battery Electric-Thermal Model. *arXiv preprint arXiv:2302.13157*.

[24] A. Mousaei, M. B. Bannae Sharifian and N. Rostami, "An Improved Fuzzy Logic Control Strategy of an Optimized Linear Induction Motor Using Super Twisting Sliding Mode Controller," *2022 13th Power Electronics, Drive Systems, and Technologies Conference (PEDSTC)*, Tehran, Iran, Islamic Republic of, 2022, pp. 1-5, doi: 10.1109/PEDSTC53976.2022.9767465.

[25] Lee, G., Min, A., and Hong, J.P., "Optimal Shape Design of Rotor Slot in Squirrel-Cage Induction Motor Considering Torque Characteristic", IEEE Transactions on Magnetics, Vol. 49, No. 5, pp. 2197-2200, 2013.

[26] Arash Mousaei, Nasim Bahari, Guo Mieho. Artificial Neural Networks (ANN) of Proposed Linear Induction Motor with Hybrid Secondary (HLIM) Considering the End Effect. American Journal of Electrical and Computer Engineering. Vol. 5, No. 1, 2021, pp. 32-39.
doi: 10.11648/j.ajece.20210501.15.

[27] Vishnu Murthy, K.M., "Computer-Aided Design of Electrical Machines", BS Publications, Hyderabad, 2008, ISBN:978-81-7800-146-3.

[28] Saygın, A., Ocak, A., Dalcalı, A., Çelik, E., "Optimum Rotor Design of Small PM BLDC Motor Based on High-Efficiency Criteria", ARPN Journal of Engineering and Applied Sciences, Vol. 10, No. 19, pp. 9127-9132, 2015.

[29] Dalcalı, A., Ocak, C., "Effect of Different Magnet Materials on The Performance of Surface Mounted Direct Drive PMSM", Journal of Awareness, 3, pp. 217-224, 2018.

[30] Sundaram, M., Mohanraj, M., Varunraj, P., Kumar, T.D., Sharma, S., "FEA Based Electromagnetic Analysis of Induction Motor Rotor Bars With Improved Starting Torque For Traction Applications", Automatic Control, Mechatronics and Industrial Engineering, pp. 103-110, Taylor&Francis Group, London, ISBN:987-1-138-60427-8.

[31] Popescu, M., Goss, J., Staton, D.A., Hawkins, D., Chong, Y.C., Boglietti, A., "Electrical Vehicles—Practical Solutions for Power Traction Motor Systems", IEEE Transactions on Industry

Applications, Vol. 54, No. 3, pp. 2751-2762, May/June 2018.

[32] Ünlükaya, E., Yetgin, A.G., Çanakoğlu, A.I., and Turan, M., "Effect of Rotor Slot Shapes on Induction Motor Performance", Symposium on Electrical-Electronic-Computer and Biomedical Engineering, pp.168-172, Bursa, November 2014.

[33] Mousaei, A., & Mohammadabadi, S. A. (2023). Optimization of a three-phase Induction Motor for Electric Vehicles Based on Hook-Jews Optimization Method. *arXiv preprint arXiv:2302.14805*.

[34] Mousaei, A., Gheisarnejad, M., & Khooban, M. H. (2023). Challenges and Opportunities of FACTS Devices Interacting with Electric Vehicles in Distribution Networks: A Technological Review.

[35] Mousaei, A., Gheisarnejad, M., & Khooban, M. H. (2023). Robust Sliding Mode Control for Two-Wheel Robot Without Kinematic Equations.

[36] Brush E.F., Cowie, J.G., Peters, D.T., and Van Son, D.J., "Die-cast Copper Motor Rotors: Motor Test Results, Copper Compared to Aluminum", Energy Efficiency in Motor Driven Systems, pp. 136-143, 2003.